



Planetary Data System

Information Model and DDWG Updates

MC Face-to-Face
UCLA
Los Angeles, California
August 26-27, 2014

Topics

- Schedule and Status of the PDS4 Information Model (IM) with respect to PDS4 Build 5a
- Data Design Working Group and Task Statuses
- Other work

Schedule

- Release of IM V1.3.0.0 for Build 5a for unit testing - Pending
- Release IM V1.3.0.0 for Build 5a for Model Integration and Testing on 9/8/14
- Release IM V1.3.0.0 for Build 5a for System Integration and Testing on 9/29/14
- Supporting requirements for LADEE, MAVEN, InSight, Osiris-Rex, *BepiColumbo*, *ExoMars*, *Juice*

IM Status

- Common model remains stable.
 - The majority of changes involve schematron rules and new permissible values.
- The release process for V1.3.0.0 is on schedule.
- Key discipline dictionaries are being test ingested for compliance.
 - cartography and geometry
- Improving user access by leveraging IM for explicit query models for specific product types
- International community is very engaged.

Release Process for the IM

- Update IM with approved change requests
- Generate documents from the IM
- Perform difference and regression tests
- Release for Unit Testing
 - Fix bugs found during Unit Testing
 - Nodes are requested to participate
- Release for Model Integration and Testing (MI&T)
 - Fix bugs found during MI&T
- Release for System Integration and Testing

CCB approved changes

Candidate changes for V1.3.0.0 of the IM

- CCB-32 - Add "ISIS2 History Label" to Header.parsing_standard_id
- CCB-56 - Align DSV References in IM and SR
- CCB-60 - Add "Calibration" to the Target_Identification <type> enumerated list.
- CCB-62 - Add standard values for units of current
- CCB-71 - Inappropriate permissible value lists for <Special_Constants> attributes
- CCB-73 - Add "Microsoft Excel" as a document_standard_id enumerated value in Document_File

International Interactions

- IPDA/PSA (S. Martinez) is reviewing the first use case for mission science data collection search. (Metadata Consistency Team)
- Regularly respond to PDS4 related questions from IPDA/PSA
- IPDA/PSA (D. Heather) has agreed to review PDS4 Data Standards documentation.
- Will forward Version 1.3.0.0 documents to selected IPDA members during Model Integration and Test for review and testing.

Documents¹

Standards Documents (in preparation for Build 5a)

- Information Model Specification – Version 1.3.0.0
- XML Schemas – Version 1.3.0.0
- Data Dictionary – Version 1.3.0.0
- PDS4 Example Products - Version 1.3.0.0
- Standards Reference - Version 1.3.0

Support Documents (planned reviews and updates)

- DD Tutorial
- Glossary
- Concepts Document
- Data Provider's Handbook
- PAG (Proposers Archive Guide)

¹Posted to <http://pds.nasa.gov/pds4>

DDWG Status and Tasks

- The DDWG now meets for one hour¹ teleconferences on alternate weeks
- Agenda
 - Research, additional study, prototyping, and proof-of-concepts
- Discipline Team Tasks
 - Geometry (focus: flyby and orbital missions)
 - Cartography
 - NSSDC/PDS Interface
 - Metadata Consistency
 - Document Improvement

¹*Has been known to exceed the allocated time*

Task - Geometry

Team lead and members: E. Guinness, M. Gordon, A. Raugh, T. Farnham, C. Isbell, S. McLaughlin, B. Semenov , C. Acton, E. Rye, S. Hughes

Short Description: Capture geometry requirements from across the disciplines, obtain a consensus model, and write a geometry dictionary. Currently focusing on classes for flyby/orbital and landed missions.

Goals: Cross-discipline Geometry Model and Dictionary

Schedule (Major milestones):

Done - Requirements captured in white paper

Sep – Next version of Orbital and Draft version of Landed dictionary

Nov– Release operational version of Geometry dictionary

Task - Cartography

Team leads: E. Rye, C. Isbell

Status: Already discussed

Task - NSSDCA/PDS4 Interface

Team lead and members: McLaughlin, Bell, McCaslin, Kodis, Hughes, Hardman, Guinness, Huber, Isbell, King, Raugh

Short Description: Develop an interface for transferring PDS4 data to the NSSDCA deep archive

Goals

- Automate deliveries; eliminate multiple ingests of products
- Provide the ability to return basic products, collections, and bundles from the deep archive; report on those entities
- Perform data integrity checks

Schedule (Major milestones)

Jul - Completed a draft of the NSSDCA's logical database model for ingesting PDS4

Aug - Finalize contents of the Submission Interface Package

Next 4-6 Months: Develop and begin testing interface; Develop new data model and ingest system."

Task – Metadata Consistency

Team lead and members: S. Hughes, J. Mafi, E. Guinness, S. Hardman, M. Martin, A. Raugh, M. Gordon

Short Description: Define explicit set(s) of attributes to be used as search parameters for products within the PDS4 archive.

Note: An implicit set of search parameters already exists, the required attributes for Product_Observational.

Goals: See next slide.

Schedule (Major milestones):

Past 2 months – Identified the first use case and have a draft consensus on the set of attributes to be considered as search parameters.

Next 2 months – Test the first use case and start development and testing of a few other use cases.

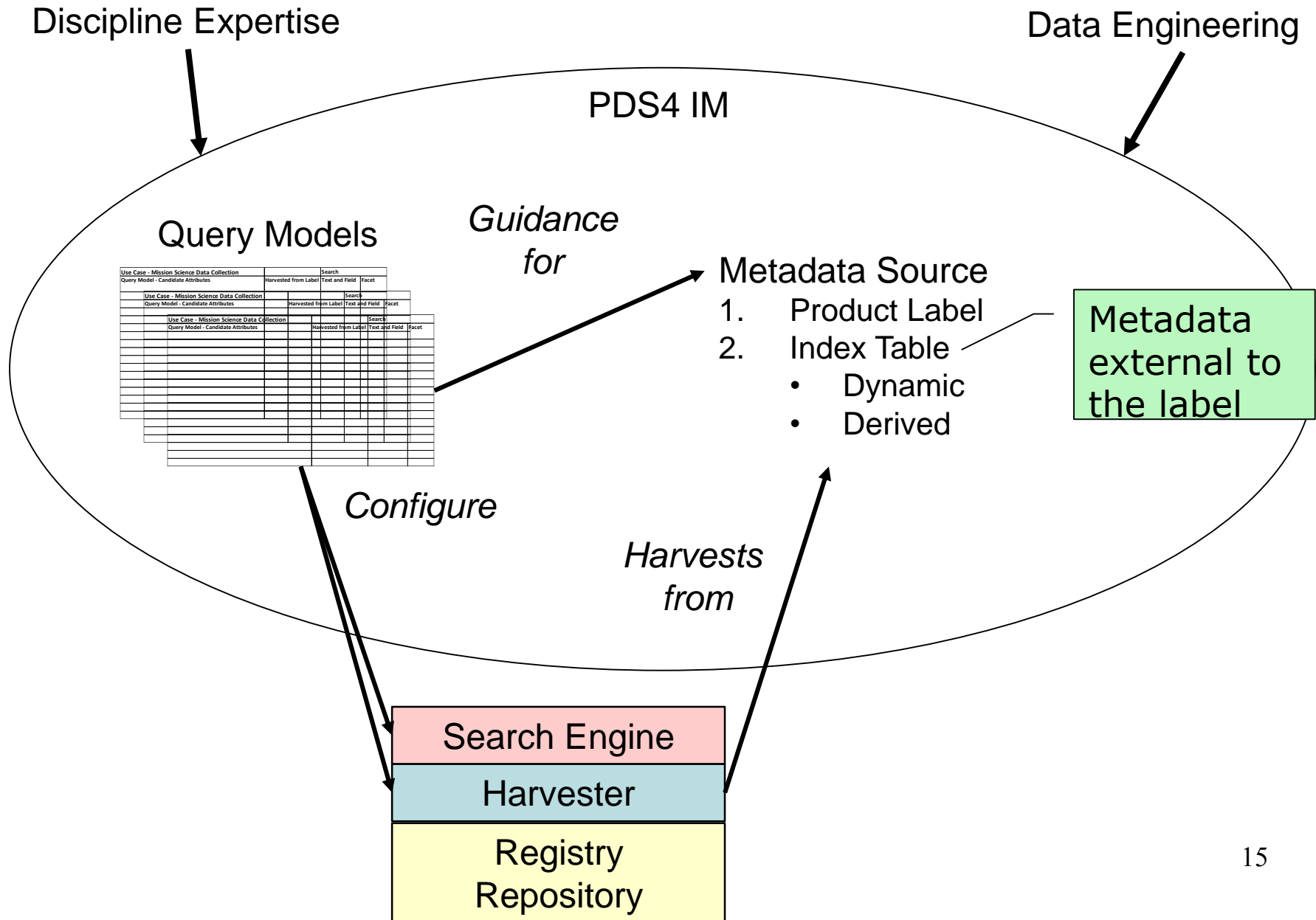
In 6 months – Complete task with a query model template, interfaces, guidelines, and processes.

Metadata Consistency

- What is the problem?
 - A single query model¹ is not appropriate for PDS4.
- What solution was planned for this problem?
 - The product-centric PDS4 architecture allows custom query models to be defined for sets of one or more product types.
- Who is responsible?
 - Definition - Node discipline experts
 - Validation and Processing - PDS4 Infrastructure
- What are the goals?
 - Define the PDS4 “query model” template.
 - Develop and test use cases for selected product types
 - Develop recommendations for how the query models are:
 - designed and maintained
 - used to configure system components
 - used to guide data providers
 - used in product label validation.

¹ Query Model - A set of attributes to be used as search parameters for products within a defined set of products.

Role of the Query Model



Task - Document Improvement - 1

Team members: D. Simpson, M. Gordon, R. Joyner

Short Description: Address issues associated with the PDS4 Data Standards documents.

Goals: Consistent data standard documents and information model.

Schedule (continuing work):

- Support Documents – Planned Reviews and Updates
 - DD Tutorial (needs rewrite to accommodate new terminology)
 - Glossary and List of Acronyms and Abbreviations (review)
 - Concepts Document (may be OK as high-level doc as is)
 - Data Provider's Handbook (needs consistency review)
 - PAG (Ron took this last spring; I don't know status)

Task- Document Improvement - 2

Schedule (continuing work):

- Receive comments on SR from IPDA (overdue)
- Follow through with revisions of enumerated values and related issues in the Standards Reference (ongoing)
- Documentation is becoming more difficult to maintain as more detail appears in Schematron files

Information Model Extracts

- Requests have been made from projects for extracts of the contents of the PDS4 Information Model in other machine languages.
 - JSON¹ – APPS (AMMOS-PDS Pipeline Service)
 - SKOS² and OWL⁵– Linked Open Data³ Project (Bernd Ritschel – GFZ Potsdam)
 - RDF⁴ – Various requesters

¹ JavaScript Object Notation (JSON) is a lightweight data-interchange format.

² Simple Knowledge Organization (SKOS) is a common data model for sharing and linking knowledge organization systems via the Web.

³ Linked Open Data (LOD) is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found, and links made between related resources.

⁴ Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax notations and data serialization formats.

⁵ The Web Ontology Language (OWL) is a family of knowledge representation languages or ontology languages for authoring ontologies or knowledge bases.

Papers and Presentations

- Paper accepted for E-SCIENCE 2014: 10th IEEE International Conference on e-Science; *A Scalable Planetary Science Information Architecture for Big Science Data*.
- D. Crichton and S. Hughes briefed a NASA OCIO WG on Information Architecture
 - The WG was enthusiastic about the model-driven approach of PDS4.

Summary

- The “common” part of the PDS4 Information Model remains stable.
 - Most changes are the type expected, for example new permissible values and schematron fixes.

“One of the things I'm liking about using the new schemas is that they do present this very complete checklist of simple questions that can be asked and answered.”

- Good progress is being made on discipline dictionaries.
 - Special teams focused on one issue are effective.
- New focus on IM's role in supporting user access.
- The international community is engaged.

Questions and Answers

Backup

Query Model – Part 1

Use Case - Mission Science Data Collection		Search	
Query Model - Candidate Attributes	Harvested from Label	Text and Field	Facet
Identification_Area.logical_identifier	Yes	Yes	
Identification_Area.version_id	Yes	Yes	
Identification_Area.product_class=Product_Collection	Yes	Yes	Yes
Identification_Area.title	Yes	Yes	
Collection.collection_type=Data	Yes	Yes	Yes
Collection.description	Yes	Yes	
Citation_Information.description	Yes	Yes	
Citation_Information.keyword	Yes	Yes	
Investigation_Area.type=Mission	Yes	Yes	Yes
Investigation_Area.name	Yes	Yes	
Primary_Result_Summary.processing_level	Yes	Yes	Yes
Primary_Result_Summary.purpose=Science	Yes	Yes	Yes

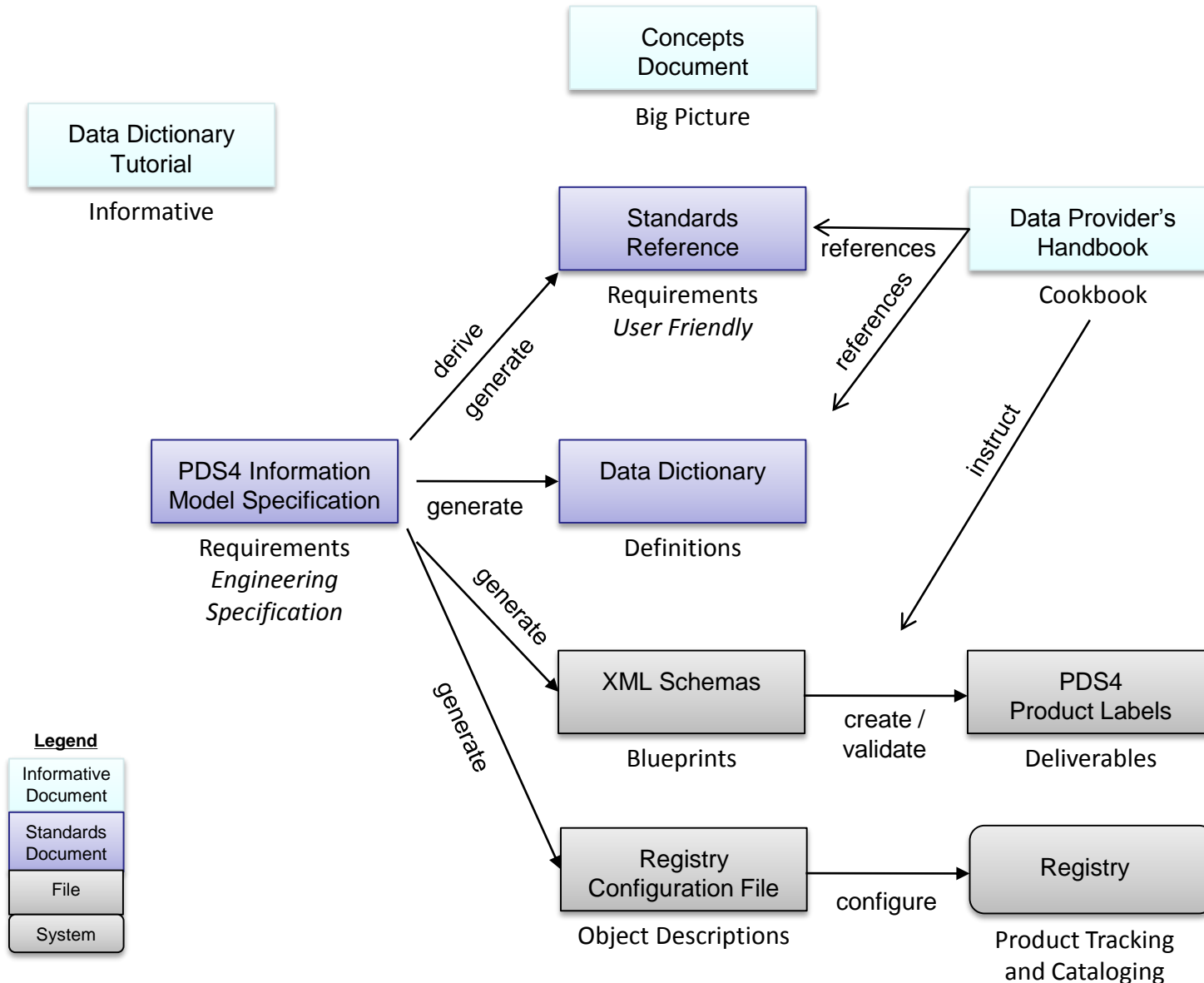
Query Model – Part 2

Use Case - Mission Science Data Collection		Search	
Query Model - Candidate Attributes	Harvested from Label	Text and Field	Facet
Science_Facets.domain	Yes	Yes	Yes
Science_Facets.wavelength_range	Yes	Yes	Yes
Science_Facets.discipline_name	Yes	Yes	Yes
Science_Facets.facet1	Yes	Yes	
Science_Facets.facet2	Yes	Yes	
Primary_Result_Summary.description	Yes	Yes	
Observing_System.name	Yes	Yes	
Observing_System_Component.name	Yes	Yes	
Observing_System_Component.type (type=Instrument)	Yes	Yes	
Time_Coordinates.start_date_time	Yes	Yes	
Time_Coordinates.stop_date_time	Yes	Yes	
Target_Identification.name	Yes	Yes	
Target_Identification.type	Yes	Yes	Yes
Target_Identification.description	Yes	Yes	
Product_Context.Instrument.type	Yes	Yes	Yes
system.archive_status	No	Yes	

Definition and Questions

- Query Model – A subset of attributes from the PDS4 dictionaries, either common or local, to be used as search parameters.
 - Used to configure the harvester and search engine
 - Can be used as guidance to data providers
 - Are mapped to registry slots
 - A select few are implemented as facets.
- Facet – A facet corresponds to an attribute in the PDS4 Information Model. For example the attribute <processing_level> could be mapped to a facet. In the PDS4 search engine a faceted classification system allows users to explore products by applying multiple filters
- Open questions
 - How many query models are needed?
 - Current general search use cases are:
 - Product_Observational
 - Product_Collection (mission science data collection)
 - Product_Document
 - How do they guide the data provider?
 - Paragraph in Standards Reference
 - Schematron Rules

PDS4 Documents and their Relationships



Testing

- Oxygen is used to test the generated XML Schemas and Schematron files for validity and well-formedness.
- File Difference (Ultra-Edit Compare) is used to identify differences between current and prior versions of generated files after updates.
 - XML Schema, Schematron, Information Model Specification, ISO/IEC-11179 Data Dictionary files.
- Regression tests are performed against a suite of example products.
- Regression tests are performed against selected node bundles (Imaging, Atmospheres)

Acknowledgements★

Ed Bell

Richard Chen

Dan Crichton

Amy Culver

Patty Garcia

Ed Grayzeck

Ed Guinness

Mitch Gordon

Sean Hardman

Lyle Huber

Steve Hughes

Chris Isbell

Steve Joy

Ronald Joyner

Debra Kazden

Todd King

John Kodis

Joe Mafi

Mike Martin

Thomas Morgan

Lynn Neakrase

Paul Ramirez

Anne Raugh

Shannon Rees

Mark Rose

Matias Roybal

Elizabeth Rye

Boris Semenov

Dick Simpson

Susie Slavney

Dillon White

Peter Allan

David Heather

Michel Gangloff

Santa Martinez

Thomas Roatsch

Alain Sarkissian

★ Anyone who sat through a DDWG 2-hour telecon or provided useful input.